



Do driver training programs reduce crashes and traffic violations? – A critical examination of the literature

Raymond C. Peck

R.C. Peck & Associates, Folsom, CA, USA

ARTICLE INFO

Article history:

Received 20 October 2010

Received in revised form 30 November 2010

Accepted 30 November 2010

Keywords:

Novice drivers

Young drivers

Driver training

Driver education

Accident countermeasures

Driver training evaluation designs

ABSTRACT

This paper reviews the evaluation literature on the effectiveness of classroom and behind-the-wheel driver training. The primary focus is on North America programs as originally taught in high schools but now also by private instructors. Studies from the United Kingdom, Australia, New Zealand and Scandinavia are also included.

By far the most rigorous study to date was the experimental study in DeKalb, Georgia, U.S.A. This study used a randomized design including a control group and a very large sample size to provide reasonable statistical precision. I reexamine the DeKalb data in detail and conclude that the study did show evidence of small short-term crash and violation reductions per licensed driver. However, when the accelerated licensure caused by the training is allowed to influence the crash and violation counts, there is evidence of a net increase in crashes.

The other studies reviewed present a mixed picture but the better designed quasi-experimental evaluations usually showed no effects on crash rates but almost all suffer from inadequate sample size. I show that as many as 35,000 drivers would be required in a two group design to reliably detect a 10% reduction in crash rates.

The advent of GDL laws in North America and other countries has largely remedied the concern over accelerated licensure of high risk teenage drivers by delaying the progress to full licensure. Conventional driver training programs in the U.S. (30 h classroom and 6 h on-the-road) probably reduce per licensed driver crash rates by as little as 5% over the first 6–12 months of driving. The possibility of an effect closer to 0 cannot be dismissed.

Some GDLs contain an incentive for applicants to complete an advanced driver training program in return for shortening the provisional period of the GDL. The results of Canadian studies indicate that any effects of the driver training component are not sufficient to offset the increase in accidents due to increased exposure.

There is no evidence or reason to believe that merely lengthening the number of hours on the road will increase effectiveness. Programs directed toward attitude change and risk taking better address the underlying cause of the elevated crash risk of young drivers but these behaviors are notoriously resistant to modification in young people.

© 2011 International Association of Traffic and Safety Sciences. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Many years ago most would have accepted as axiomatic the premise that pre-license driver training leads to increased driving skill and fewer crashes. This assumption, in fact, led to the creation of the professional driving school industry in the United States during the 1930s. Driver-training (classroom and on-the-road) ultimately became inculcated into the curriculum of many high schools and by 1960, many U.S. states required teenage drivers to complete a certified classroom and behind the-school program before receiving their original driver's license. The required training usually consisted of 30 h of classroom education and 6 h of on-the-road instruction (1).

During this period, a number of rather extravagant claims were made by the driver training industry, sometimes in concert with insurance companies, claiming that driver training programs produced large reductions in young driver crash rates. Some insurance companies offered discounts to teenage drivers who had completed driver training (2–4).

A casual inspection of the data and the studies cited to support these effectiveness claims indicated them to be void of any validity. In all cases, enrollment was voluntary and in some cases there was additional selectivity by school personnel based on academic performance (4). Subsequent research confirmed that self-selected volunteers had much more favorable characteristics than did comparison groups of non-trained students. Thus, any differences on subsequent record were confounded by variables such as socioeconomic status, gender, social adjustment, grade-point average and

E-mail address: Homepeck@aol.com.

Table 1
Unadjusted crash and violation means by gender and driver training status (4).

Variable	Males			Females		
	Trained (N = 3978)	Not trained (N = 2445)	Sig.	Trained (N = 2858)	Not trained (N = 1907)	Sig.
Total crashes – 1 year	.151	.176	$P < .05$.085	.104	$P < .05$
Fatals and injury crashes – 1-year	.044	.057	$P < .05$.018	.035	$P < .05$
At fault crashes – 1-year	.025	.032	N.S.	.009	.022	$P < .05$
Single vehicle crashes – 1-year	.012	.012	N.S.	.003	.010	$P < .05$
Traffic citations – 1-year	.567	.819	$P < .01$.140	.198	$P < .05$
Traffic citations – 3-years	2.11	2.70	$P < .01$.543	.694	$P < .01$

intelligence (5–7). No attempt was made in these early studies to adjust subsequent differences in crash rates for the aforementioned biases.

Since the late 1960s, three types of research designs have been used to estimate the causal effect of driver education training on subsequent crash and traffic violation rates:

- (1) Retrospective or prospective quasi-experimental designs comparing trained and untrained drivers after adjusting for pre-existing differences through matching, stratification or analysis of covariance. These designs are subject to model specification errors and confounding by omitted variables.
- (2) Randomized control trials (RCT) in which assignment to trained groups or a non-trained control is random. These designs are considered the “gold standard” for establishing cause and effect relationships but are difficult to execute successfully due to logistic, ethical and legal constraints. They are also subject to experimental artifacts.
- (3) Ecological designs in which the quantities are aggregate measures such as the number of drivers licensed by age, number trained and rate of crashes over periods of time in different geographical regions, such as states. These designs are often subject to serious confounding, endogeneity bias, and problems in generalizing ecological relationships to the behavior of the entities of interest – i.e., individual drivers.

This paper reexamines the key research evidence concerning the effects of driver training on per capita and per licensee crash rates and discusses how the implementation of GDL laws in many jurisdictions has altered the policy implication of past driver training evaluations.

2. An early quasi-experiment

The first quasi-experimental driver training study to formally model the non-random assignment process using multivariate methods on a large representative sample of novice drivers was the California study by Harrington (4). Harrington performed a longitudinal analysis of the first four years of driving of 13,915 novice drivers aged 16–17 at the time of initial licensing in 1963.

At the time of sample selection, driver training in California was voluntary. The unique relevance of Harrington's study was the large number of biographical, socio-economic and social-adjustment variables collected and the use of these variables in identifying differences between students volunteering for driver training and those who did not. Included in the data set were variables collected from each driver's school record, including grade-point average, citizenship ratings, truancies, I.Q., achievement test score, home status, and driver training status. Additional data were collected through a mailed questionnaire and, for a small subset, through personal interviews.

Using correlational and multiple regression techniques, Harrington first identified those variables on which the trained and non-trained group differed. For males, significant univariate differences were found on 50 variables. For females, significant differences were observed on 29 variables. A stepwise multiple regression analysis produced multiple R s of .42 for males and .35 for females in differentiating the trained group from the non-trained group. The untrained group had significantly poorer scores on the stronger crash predictors (e.g., grade-point average, citizenship ratings, school-attendance and achievement tests) in a direction that was associated with increased crash and traffic violation rates.

Summarized in Tables 1 and 2 are driver record comparisons (mean frequency) on the key dependent variables prior and subsequent to analysis of covariance adjustment.

The results and implications of these results are clear. Prior to adjustment for self-selection volunteer bias, driver training appears to have had a significant beneficial effect on crashes and violations for both males and females. But after adjustment, none of the differences approached significance for males. For females, however, all of the differences except one (total crashes) still showed a significant ($P \leq .05$) or suggestive ($P \leq .10$) effect in favor of training. Thus, there was clear evidence of a training \times gender interaction in which training effects were moderated by gender.

It is reasonable to question how these results generalize to the present given the age of the study even if one accepts the interaction effect as representing a causal effect of training. There have been major changes in the role of gender in driving and crash involvement over the past 40 years. This could explain why the gender \times training interaction did not replicate in a later experimentally controlled study by Stock et al. (8).

Table 2
Bias-adjusted crash and violation means by gender and driver training status (4).

Variable	Males			Females		
	Trained (N = 3978)	Not trained (N = 2445)	Sig.	Trained (N = 2858)	Not trained (N = 1907)	Sig.
Total crashes – 1 year	.162	.158	N.S.	.092	.095	N.S.
Fatals and injury crashes – 1-year	.050	.048	N.S.	.021	.032	$P < .05$
At fault crashes – 1-year	.027	.027	N.S.	.010	.020	$P < .05$
Single vehicle crashes – 1-year	.013	.011	N.S.	.004	.010	$P < .05$
Traffic citations – 1-year	.654	.673	N.S.	.154	.176	$P < .10$
Traffic citations – 3-years	2.32	2.31	N.S.	.583	.634	$P < .10$

My main purpose in referencing the Harrington study in some detail is to illustrate the problems inherent in any observational cohort study where assignment to a treatment group is voluntary. The appropriate specification of the assignment bias requires access to a covariate set representing a broad array of criterion relevant characteristics. It is essential to have very large sample sizes with multivariate data, particularly when the dependent variables have small means and variances, which is always the case with crash and traffic citation rates (9,10). The availability of school record data, in particular, was critical to the bias adjustment used in the Harrington study because of the substantial relationship between personality, lifestyle, social adjustment and crash propensity (4,9,12). Harrington concluded the paper with a recommendation that future evaluations of driver training utilize experimentally controlled random assignment designs.

3. The Dekalb County study

This study, despite some flaws, is by far the most definitive evaluation of driver training to date because of its relatively large sample size ($N = 16,000$) and random assignment design. It therefore is summarized and discussed at considerable length in this review.

DeKalb addressed the volunteer bias issue by first identifying a pool of students who intended to become licensed and who were agreeable to participating in the study. Subjects were then randomly assigned to one of the two training programs or a no-training control group while simultaneously matching them on grade-point average, gender and socioeconomic status. All three of these variables are known to be related to crash and traffic violation risk and the matching procedure provided an additional assurance that the design would be balanced on these three factors. The violation and case records were tracked and collated over a 2–4 year post-treatment period.

The components of the two training programs are described in Table 3. The PDL (pre-driver license) was designed as a “bare bones” program designed to provide the minimum training needed to pass the driver license exam. The SPC (safe performance curriculum) was characterized as a “state of the art” program designed to enhance drive competency in areas known to be critical to safe driving and crash avoidance, including hazard perception. The program evolved from a plan and series of intermediate projects beginning in 1970–1971 with an in-depth drive task analysis and culminating in the DeKalb County evaluation, 1977–1983 (8).

The control group was provided no-training by the school or program. The idea was for the control group to represent how students in DeKalb County would acquire driver training in absence of a school-based program (e.g. parents, friends, commercial schools, etc.). However, they were provided an incentive for study participation in the form of insurance discounts by passing a specially developed drive test (approximately 10% of the controls qualified for the discount) (11). The final report by Stock et al. (8) contains a series of analyses organized around three sample breakdowns:

- (1) Total assigned sample from point of random assignment.
- (2) The subset of the total sample that had at least 6 months of exposure as a licensed driver.
- (3) A subset of #2 who also completed their assigned treatment program (SPC or PDL).

(3) A subset of #2 who also completed their assigned treatment program (SPC or PDL).

The primary analysis involved #1 since this provided the only comparisons in which randomization was retained. As such, one has confidence that self-selection and confounding have been well controlled. The second and third analytic subsets were motivated by interest in estimating the actual effect of treatment assignment on the crash and violation rates of drivers subsequent to licensure and program completion. These analyses are subject to confounding since not all assigned drivers became licensed or were licensed at the same time.

In addition, 28% of those assigned to training either did not enroll or did not complete the training. Since program dropouts are not random and often have “less desirable” characteristics than those who complete programs, their exclusion introduces a potential bias that understates the true crash and violation rate of the trained group. However, their exclusion engenders a potential bias in the opposite direction. Those who do not enroll or complete training would not have received whatever benefits that these programs might provide. In short, an analysis based only on assigned subjects might not capture the actual effect of completing the training. A comprehensive evaluation therefore requires an analysis of each of the three sampling units described above as reflected in the original report by Stock et al.

A summary of the effect comparisons on crash and violation rates for the 24-month period following assignment is shown in Table 4. None of the differences between the two trained groups approached significance and neither of the two trained groups differed significantly from the controls with respect to the mean rate of crashes or violations. The only comparison directionally favorable to training involved the “bare bone” program (PDL) but the differences were extremely small and entirely consistent with sampling error.

Stock et al. also evaluated the 2-, 3-, and 4-way interactions among the factors. None of the interaction terms involving treatment were significant for crashes, including the interaction between training and gender. Thus, the results for crashes were consistent with an additive effects model.

An unfortunate limitation of Stock et al.’s analysis of the data in Table 4 is that the post assignment period of the subjects varied from 1–24 months. Since assignment was random and matched across treatment, this would not have created a bias but it does decrease the precision of the analysis. It is not clear why Stock et al. did not control this source of variance by including follow-up time as a covariate or denominator since it would have increased the statistical power for detecting differences between the treatments.

Interpretation of the above finding was further complicated by the finding that training tended to cause students to become licensed sooner and to become licensed at a greater rate. For example, 69% of the trained group was licensed after 6 months compared to 59% of the controls. After 24 months, the difference declined to 87% (PDL and SPC combined) to 84% (controls). On the average, the trained group accumulated 23 more days of licensed driving over a 24 month follow-up period than did the controls. Using a different method, Lund et al. estimated that PDL and SPC subjects were 10%–16% more likely, respectively, to have obtained a license in any given month during the post-assignment period (11).

Table 3
SPC and PDL program elements.

SPC	PDL
1. 32 h of classroom instruction and range training	1. 20 h of classroom simulator
2. 16 h of simulator instruction	2. One hour of on-the-road training Supplemented by parental training
3. 16 h of driver range instruction	
4. 3 h of evasive maneuver training	
5. 3.5 h of on-the-road training	

Table 4

Subsequent cumulative accident and violation rates by treatment group as reported by Stock et al. (1983) (1–10 quarters — all assigned subjects).

Treatment	Accident means	Violation means
SPC	.378	.977
PDL	.361	.956
Control	.364	.977

Note: all differences are non-significant.

Because of the above, the rates in Table 4 also reflect different levels of licensed exposure. The control group, in particular, had more licensed exposure than the two trained groups, and increased exposure is known to have a direct effect in increasing the risk of crash and violation involvement. Ordinarily, one would adjust the rates for this fact. However, this would be problematic for these data because the increased licensure and earlier licensing were, in essence, caused by the training and are therefore part of the net effect. Such an effect should not be surprising since the 2 trained groups had direct access to training and the knowledge and skill enhancement required to pass the Georgia knowledge and on-the-road test.

Under an “intent to treat” (ITT) paradigm, the analyses quite properly leave all subjects in the assigned group because the randomization occurred at this point. Nevertheless, students who did not enroll or complete the program did not receive whatever benefits were provided by the training. It might be argued that program dropouts are an inherent part of any program where enrollment and completion are voluntary. However, this ignores the fact that in actual practice, many states would not license teenagers who did not complete the prescribed training. This was not true of Georgia where driver training is optional.

Stock et al. addressed the above concerns by conducting 2 additional sets of analyses: one confined to licensed drivers and the other to licensed drivers who also completed the assigned training program.

In summarizing these results, I have collapsed the SPC and PDL into a single training group since Stock et al. found no difference between SPC and PDL on subsequent crash and violation rates. I have recalculated the analysis of variance significance tests to reflect a 2 group as opposed to a 3 group design.

Since some authorities have questioned Stock et al.'s use of analysis of variance (ANOVA) on highly skewed dependent variables, there needs to be a brief defense. A number of Monte Carlo studies have found fixed effects ANOVA to be highly robust to even extreme non normality as long as *N*s are large and heterogeneity of variance is not extreme (13,14). There can be distortions if the data are highly non-orthogonal and smaller *N*s are associated with larger variances. However, the differences in variances in these data were very modest.

Table 5 shows comparisons between trained and untrained drivers for the 2 sample subsets over 4 periods of time from licensure: first 6 months, second 6 months, third 6 months and fourth 6 months. For licensed drivers, the trained group had 13.1% fewer crashes during the first 6 months of license driving ($P \leq .07$). The percent reductions declined after the first 6 months and the directional reductions in favor of training did not approach significance after the first period. Across all 4 six-month periods (24 months), the trained group had

8.8% fewer crashes, which was just short of statistical significance ($P > .10$).

Part II of Table 5 limits comparisons to licensed drivers who actually completed the training. These comparisons show more evidence of a positive effect in favor of training. The 16.4% crash reduction at period 1 and the 8.3% reduction after 24 months were significant at $P \leq .02$ and $P \leq .03$, respectively. Unfortunately, these comparisons are confounded by self-selection bias stemming from the likely characteristics of program dropouts. Stock et al. observed that students with low GPAs were more likely to not enroll or drop out and to also have higher crash rates. Any bias would have at least partially been controlled since GPA was a design factor in the analysis of variance. It is not clear why Stock et al. did not model the assignment bias by comparing dropouts with program completers on all available covariates and adjusting the results through analysis of covariance or multiple regression procedures.

Despite the above limitations, some guarded interpretation of the results for the program completions are offered here. The first thing to note is that the profile of the differences between I and II are quite similar. The largest difference in effect size occurred in period 1 (13.1% vs. 16.4%). This is not a large difference and the other periods are remarkably similar. One would ordinarily expect a strong bias due to attitudinal and lifestyle difference to exert an effect over a longer time period. The fact that the effect size is slightly larger for program completers is also what one would expect if the training had some impact in reducing crashes.

Presented in Table 6 is a summary of the effects on traffic violations, using the same format as in Table 5. The effects on violations are more pronounced and longer lasting than for crashes. For licensed drivers, the violation reductions directionally favored the trained group in all 4 periods and are statistically significant in periods 1, and 2 at $P \leq .05$ and for all 24 months combined at $P \leq .06$.

For those who completed the program, the effect for period 1 (19.4% reduction, $P \leq .01$) is more pronounced than for licensed only drivers but the pattern for the other periods is very similar. In fact, the effect size in the second 6 months is actually lower for licensed program completers than for the total licensed group. Both sets suggest a positive effect lasting 12–18 months.

The fact that treatment differences on violations are more significant than on crashes is consistent with prior research (9) and is readily explainable by the fact that violations more directly reflect driving behavior than do crashes. As with crashes, the effect pattern over time seems more consistent with a real treatment effect that is dissipating over time than with differences in the characteristics of the groups. Of course, there could be a combination of a bias and a valid training effect.

Table 5

Analysis of DeKalb crash rate differences by time period for licensed drivers and licensed drivers who completed training.

Group	I. licensed drivers by follow-up period				
	1st 6 months (<i>N</i> = 12,928)	2nd 6 months (<i>N</i> = 11,607)	3rd 6 months (<i>N</i> = 9460)	4th 6 months (<i>N</i> = 7468)	2-year aggregate (<i>N</i> = 7468)
A. trained	.106	.098	.084	.095	.383
B. control	.122	.101	.093	.095	.420
% difference (B – A/B)	– 13.1%	– 3.0%	– 9.7%	0	– 8.8%
Significance level	$P < .07$	N.S.	N.S.	N.S.	N.S. ($p = .11$).
Group	II. licensed and completed program by follow-up period				
	1st 6 months (<i>N</i> = 11,055)	2nd 6 months (<i>N</i> = 9950)	3rd 6 months (<i>N</i> = 8094)	4th 6 months (<i>N</i> = 6368)	2-year aggregate (<i>N</i> = 6368)
A. trained	.102	.097	.083	.096	.385
B. control	.122	.101	.093	.095	.420
% difference	– 16.4%	– 4.0%	– 10.8%	+ 1.1%	– 8.3%
Significance level	$P < .02$	N.S.	N.S.	N.S.	$P < .03$

Table 6

Analyses of DeKalb traffic violation-rate differences over time period for licensed drivers and licensed drivers who completed program.

Group	I. licensed drivers by follow-up period				
	1st 6 months (N = 12,928)	2nd 6 months (N = 11,607)	3 rd 6 months (N = 9460)	4th 6 months (N = 7468)	2-year aggregate (N = 7468)
A. trained	.151	.159	.194	.200	.736
B. control	.175	.188	.211	.214	.815
% change (B – A/B)	– 13.7%	– 15.4%	– 8.1%	– 6.5%	– 9.7%
Significance level	$P < .05$	$P \leq .03$	N.S.	N.S.	$p \leq .06$
Group	II. licensed and completed program by follow-up period				
	1st 6 months (N = 11,055)	2nd 6 months (N = 9950)	3 rd 6 months (N = 8094)	4th 6 months (N = 6368)	2-year aggregate (N = 6368)
A. trained	.141	.166	.190	.205	.723
B. control	.175	.188	.211	.214	.815
% change (B – A/B)	– 19.4%	– 11.7%	– 9.9%	– 4.2%	– 11.3%
Significance level	$P < .01$	$P \leq .09$	N.S.($P \sim .20$)	N.S.	$P \leq .03$

3.1. Reanalysis of the DeKalb data

The original data collected in the DeKalb study have been reanalyzed by other investigators but only the analysis by Lund et al. has been published in a peer-review journal. These investigators (11) confined their analysis to the random assignment component of the study and used a different start date in defining the beginning of the post-treatment interval for counting crashes and traffic violations. Instead of the actual assignment date, Lund et al. compared the 3 groups from the point of their 16th birth date. In contrast to the original analysis of variance of mean rates, the Cox proportional hazard technique was used in measuring time to licensure, time to first crash and time to first traffic violation. This approach offers certain advantages over analysis of variance since it allows for different exposure periods and entry dates among subjects and, when the failure events are rare, provides a more informative and sensitive measure of treatment effects. A disadvantage is that statistical power is diminished if there are substantial numbers of crash and violation repeaters. The proportion of crash repeaters in the DeKalb data was small but the proportion of violation repeaters was substantial.

Lund et al. found that the SPC was associated with a significant increase in the proportion of crash and traffic violators compared to the control group ($P \leq .01$). The estimated increase in the hazard rates for crashes and violations were 8% and 11%. The crash risk differences between the PDL and control group were not significant ($P \geq .05$). Consistent with Stock et al., both training programs significantly decreased the time period to licensure. For the SPC, for any given month students were licensed 16% sooner than the controls. For the PDL, there was a 10% acceleration in licensure rate.

In concluding their paper, Lund et al. characterized their findings as follows:

These results lead to a different conclusion than that of Stock et al., in regard to the per capita driving risk of teenagers in the DeKalb County study. Despite the presence of factors that would constrain the licensure effect of driver education, students assigned to SPC were at significantly greater hazard of crashing and of receiving traffic violations than were comparable to the control students. There was no evidence that SPC (or PDL, for that matter) reduced the per capita likelihood of crashes or violations, even during the first six months of eligibility for licensure. Only when crashes and violations were analyzed per licensed driver did the results favor driver education [Stock et al., 1983], and as discussed earlier, this analysis does not provide a valid test of driver education.

For a number of reasons, the analyses by Lund et al. represents a stronger analysis concerning the net effects of driver training on per

capita crash and violation rates. The difference is largely attributed to the differences in the effect metrics used (proportional hazard model vs. analysis of variance), but there needs to be some qualification regarding the different findings on violations. The Cox proportional hazard model only considers the time to first incident. This contrasts with the analysis of variance of means, which examines the total number of incidents, including repeats, over the criterion period. The Cox proportional hazard model does not capture this component of post-treatment driving performance.

Some of the conclusions offered by Lund et al. require further comments, particularly their contention that any analyses based on subsets of the total sample and use of per licensed driver rates “do not provide a valid test of driver education.” Even if one accepts the premise that a per capita analysis based on the total pool of assigned drivers is the preferred unit, I would argue that an analysis based on all three units (assigned drivers, licensed drivers and licensed drivers who completed the program) provides a more comprehensive and fairer appraisal. The fact that this approach creates complexities and interpretive ambiguities requiring judgment does not mean it is invalid or unscientific. The key question is whether or not the separate analyses address different questions and hypotheses of interest. The analysis based on all assigned subjects has the advantage of representing the net effects of treatment and maintaining random assignment. But it could be argued that the populations of interest are those who actually drive. If the self-selection mechanisms are negligible, comparisons limited to licensed drivers would not be subject to substantial bias, particularly given statistical control over gender, SES and GPA.

The analysis based on students who were licensed and completed their assigned training program was motivated by a desire to estimate the actual effects of receiving the training programs. Persons assigned but who did not enroll or complete training would not receive whatever benefits were produced by the training. If there is interest in evaluating this effect, it is not directly captured in a comparison based on assigned groups. The exclusion of the dropouts is clearly a potential source of bias but there is also a potentially strong counter-bias by classifying drivers as treated who did not really receive the treatment. The exclusion becomes all the more dramatic when recognizing that it is an artifact of conducting the experiment in a state in which young novice drivers could become licensed without completing a certified driver training program. In states like California which require training, these dropouts and non-enrollees would not have been licensed and for this reason the program completion rate would have been higher. One can only speculate on the likely effects of this artifact but there is extensive literature showing that intent-to treat (ITT) designs tend to underestimate the true effect of treatments when non-compliance is substantial. Efron and Feldman (15) present approaches for adjusting for non-compliance in evaluation of drug

treatments and Porta et al. (16) conducted a systematic pooled analysis of 72 published papers that used both approaches (all randomly assigned subjects vs. the subset completing the assigned treatment). They found that the ITT paradigm tended to produce more conservative estimates of treatment effects but that decisions regarding statistical significance concurred in 85% of the studies. The authors stressed the need to identify variables associated with non-compliance so that the non-compliance mechanism can be modeled statistically and minimized. Lund et al. offer a number of reasons for speculating that the DeKalb results actually understated the extent to which driver training increases crashes and accelerates licensure. I do not find their arguments compelling and there is no acknowledgment of possible factors having an opposite effect.

Following publication of Lund et al., several unpublished analyses were conducted by NHTSA staff, culminating in a 1990 reanalysis by Davis (17). Only the report by Davis is considered in some detail here.

Davis' reanalysis was limited to the randomly assigned component of the study for the same reasons emphasized by Lund et al. But in contrast to Lund et al., he placed the beginning of the post-treatment period at the beginning of training or control assignment. He used a weighted least square procedure with a repeated measure structure representing the four one-year periods following treatment. Within each year, the dependent variable measures were binary (e.g. one or more events), but repeat events occurring in different years would be represented. (It should be noted that his analysis included two additional years of data not available at the time of the original analysis.) Davis' results are summarized below:

1. The controls had significantly fewer crashes and violations than either training group in year 1 ($P \leq .001$). The respective mean percent increase over the controls for crashes and violations was 16% and 8.7%.
2. None of the differences between the controls and trained groups in years 2–4 were significant.
3. None of the crash and violation rate differences between the SPC and PDL were significant in any of the years. However, the SPC had directionally more crashes and violations than the PDL in each of the 4 years.

The above results are much more consistent with Lund et al.'s findings than with Stock et al.'s total assigned group analysis and are similarly explainable by the increased licensed driving exposure of the trained groups.

The strengths and weaknesses of Davis' analyses are generally similar to those of Lund et al. The Davis analysis has the advantage of additional years of data and also of partially accounting the incidence of crash and violation repeaters. However, like Lund et al., it makes no attempt to control for temporal exposure while licensed.

It is surprising that none of the DeKalb studies made any attempt to identify the assumed non-random factors affecting the treatment comparisons among the licensed and complete-training sub samples. The most rigorous approach to this problem would have been use of the propensity score method developed by Rosenbaum and Rubin (14,19). This model would provide a score representing the propensity to become licensed vs. not licensed and complete training vs. dropout given their scores on a vector of covariates. The resultant score can then be used to match or stratify the groups in order to control difference in non-random assignment propensity.

3.2. Implications of SPC vs. PDL findings

Comparisons between the SPC and PDL have not been emphasized in the above review but in a sense are more surprising and important than those involving comparisons with the control group. They are also "cleaner," since any self-selection factors and differences in licensing rates are much less than those involving the control group. A summary of what Stock et al. found can be stated in one sentence:

there were no significant differences between the two trained groups in any of the analyses and time periods. In contrast, the analyses by Lund et al. and Davis suggest a slight advantage in favor of the PDL. Thus, a very minimal training program that did not even meet the conventional 6-hour on the road standard used in many U.S. states did as well, if not better, than a "state of the art" program. This counterintuitive result becomes more enigmatic given the findings on the intermediate outcomes. For example, Stock et al. reported that the SPC was associated with significantly higher scores on very reliable tests of safe driving knowledge and on the road performance.

Some investigators have suggested that any competency enhancement from the SPC might have been offset by increasing drivers' confidence in taking increased risks. Indeed, there is evidence from other sources that this can occur (20). The most obvious source for such an effect might be in the SPC training module for evasive, crash maneuvers. It is not unreasonable to hypothesize that, even if competency in crash avoidance maneuvers were increased by training, novice drivers would become more confident about their ability to avoid crashes. If the increase in accident avoidance skill were less than any increase in willingness to assume risky behaviors (e.g. driving over the speed limit), the net effect might be increased crashes.

Whatever the reasons, the DeKalb findings raise serious questions about efforts to increase hours of on-the-road training beyond minimum requirements without further research support.

4. Post DeKalb studies and literature review

Following DeKalb, a number of literature reviews have been published, most notably Nichols (2), Mayhew and Simpson (21,22), Woolley (23), Roberts and Kwan (24), Lonero and Mayhew (18), Christie (20), Vernick, Li, Ogaitis, MacKenzie, Baker and Gielen (25) and Masten (26). In some cases, these literature reviews are really reviews of prior reviews rather than independent reviews of the same set of empirical studies. Most are non-methodologically oriented in that they do not critically assess research design flaws and artifacts that impose limitations on study findings and between-study compatibility. Nevertheless, conclusions reached by the reviews are consistent: there is little or no compelling evidence showing that driver training reduces the crash rate of novice drivers and that any small effects are offset by a tendency of high school driver training programs to increase licensure rates at younger ages.

This was the major conclusion reached by Mayhew and Simpson in their comprehensive 1996 analysis of 30 driver training studies (21). A later review by these same authors included four additional papers each of which concluded that driver training increased crashes by increasing licensure and crash exposure (22). It will be noted that this latter conclusion is consistent with the findings from DeKalb as interpreted by Lund et al. and Davis and their preference for a per capita crash rate metric.

The most formally systematic of the latter four reviews was by Vernick et al. of the Johns Hopkins School of Health (25). These authors identified 27 studies but limited the selected studies to those considered to be methodologically sound. Only nine studies met their criteria for inclusion. Five of the studies were ecological studies using aggregate parameters (e.g. state or community-wide crash rates, number of trained teenagers, etc.) and four were randomized controlled trials. But 3 of the 4 randomized studies were all based on the DeKalb data so they are not really independent studies. The fourth randomized study was an Australian study by Strang et al. (27). This study was well done but involved only 742 drivers assigned to 1 of 3 treatments. Thus, the results are essentially non-informative because the statistical power for detecting even substantial effects on crash rate would be extremely low (see Section 5).

Not included among the randomized studies is the California study by Dreyer and Janke (28). This study compared a group receiving a

specifically developed drive range training program ($N = 1139$) with a traditional on-the-road training group ($N = 918$). An analysis of post-one year driver records showed that the range group had 33% fewer crashes than the on-the-road group ($P \leq .05$).

The study was excluded from the set of studies selected by Vernick et al. because it did not include a “no treat” control condition. However, this decision seems disingenuous given the fact that prior evidence and Vernick et al.’s own conclusions indicated that conventional on-the-road training programs have little or no effect on crash rates. If one accepts this conclusion, comparisons between an experimental program and standard program can be directly generalized to comparisons with a pure control condition.

It is also difficult to reconcile the uncritical acceptance of the five ecological studies with the requirement for randomized assignments used for selection of the studies based on individual driver assignment. Ecological studies are not only subject to the so-called ecological fallacy, they are subject to confounding and problems in disentangling the direction of causation (endogeneity bias) (29).

The five ecological studies actually produced very disparate effects on the magnitude of the associations although they did show that states with mandatory driver training laws tend to license a higher proportion of pre-18 year olds. However, none of these ecological studies employed an intervention time series design, making it difficult to establish the direction of causation.

The two cited ecological studies by Levy appear to show that when licensure rate and minimum licensing age are controlled, that driver education is associated with decreased fatal crash rates (30). This finding tends to support the conclusions suggested by Stock et al., namely, that driver training has a short term positive effect on licensed drivers that is offset by earlier licensure.

A recent report by Lonero and Mayhew contains a comprehensive review of the driver education/training literature, including studies published after the DeKalb studies (18). With the exception of the California study by Dreyer and Janke, which was discussed earlier and a study in England by Wynne-Jones (31), none of these studies utilized random assignment and all suffer from other methodological limitations, such as inadequate sample size and use of self-report driving record data.

The study by Wynne-Jones showed no evidence of an effect on self and police reported crash rates but the sample size (561 trained and 227 controls) was far too small to reliably detect an effect even if one existed. The cited study by Masten and Chapman maintained randomization but was limited to the classroom phase and did not include driver record measures (32).

A Swedish study by Gregersen (33) was described as “approximately random” by Lonero et al. but it did involve some non random geographical differences in the composition of the treated and control group (24). This study was also limited to self-reported measures including crashes. This could explain the counterintuitive finding of a significant increase in crash rate (per kilometer) due to training in the first year and a decrease in crash rate for the trained group in the second year. Such a paradoxical effect pattern is more suggestive of a self reporting artifact than a true effect.

Lonero and Mayhew also reviewed 7 quasi-experimental studies published subsequent to DeKalb and listed in Table 1 of their report. These studies were done on driver training programs in the United Kingdom, Australia, Pennsylvania, Manitoba, British Columbia, Ontario and Texas. The British Columbia study is briefly described in Section 6 of this paper. The other six countries’ studies contain at least two of the following methodological limitations: uncontrolled self-selection bias, inadequate sample size and use of self-report crash and traffic violation data. Some claim evidence of crash reduction, some no effects and others of increased crashes. But because of the limitations, they are essentially non-informative as to the effects of driver training on crash rates and for that reason are not summarized further here. The interested reader is referred to Lonero and Mayhew (18).

5. The sample size problem

The question of sample size has emerged frequently as an issue in the above discussion and is not adequately appreciated by past investigators and reviewers beyond vague acknowledgements of the need for large sample sizes to detect effects on infrequent rare events. But this leaves unanswered the question of “how large”? Table 7 shows the sample sizes needed for reliably detecting given percentage reductions in crash rates assuming that the training program reduced crashes by at least the amounts shown. These data are based on the means and variance of the crash rates of California 16–17 year olds for the first 6 and 12 months of driving (34).

The per group sample sizes required to detect a 10% reduction in 12-month crash rates with 80% confidence are 17,500 – a total of 35,000 subjects. For a 5% reduction, the N increases to 70,000 per group. These sample sizes substantially exceed even those of the DeKalb study. A 20% effect could be reliably detected with much smaller N s but even here the sample sizes greatly exceed those used in the great majority of the reviewed studies. These results have several implications. First, the finding of no significant effects in many past studies has little informative value given the frequently inadequate sample sizes of the studies. Second, even in large sample studies that are methodologically sound, small non-null positive effects cannot be reliably detected. No driver training study that has been done could reliably have detected a 5% crash reduction.

One approach to mitigating the sample size problem is to pool effect sizes across numerous studies utilizing Meta analysis. However, the disparate rigor and hypotheses being tested in the studies present serious challenges to performing a meaningful Meta analysis. One such study from Norway (35) was referenced by Lonero and Mayhew but is not yet available in English and could not be adequately assessed for this paper. In conducting a Meta analysis, incidentally, it is important to avoid the temptation of treating the various DeKalb analyses as separate studies.

6. Effect of GDL

The adoption of graduated licensing programs in North America and many other countries has rendered the issue of increased licensure rates relatively moot by increasing the time period and conditions for achieving full licensure. This effect is illustrated in Table 8 for California. The average licensure rate for the 5 years prior to GDL was 31.6% compared to 27.4% for the 7 years following GDL – a percentage decline of 13.3%.

Numerous other evaluations of GDL have also found that GDL reduces exposure during the initial provisional stage of licensing and, as a result, a reduction in crashes involving 16 and 17 year olds. This leaves us with the question of whether driver training should still be required before a driver can be granted a GDL. If one agrees that the concern over the effects of accelerated licensure have been remedied by GDL, the best evidence resides in Stock et al.’s analysis of the licensed driver results presented in Tables 5 and 6. Although some have argued that the positive effects most likely represent selection

Table 7

Per group sample sizes needed to detect specified crash reduction effects of novice driver training programs at power = .80 and $\alpha \leq .05$, two-tailed (two group design – treatment vs. control).

Assumed effect size*	Follow-up intervals	
	6 months	12 months
5%	130,000	70,000
10%	34,000	17,500
20%	8,500	4,500

* Percent difference in means of the trained and untrained group using California driver record rates for 16–17 year olds. Significance-test based on a t -test. Chi-square test of percentage difference in crash-involved driver yields similar sample sizes.

Table 8

Per capita licensure rates of California drivers aged 16–17 year olds prior and subsequent to 1998 GDL Program*.

Year	Percent of population licensed
1993	33.7
1994	33.0
1995	29.3
1996	30.9
1997	31.3
1998(GDL)	–
1999	28.2
2000	27.7
2001	28.2
2002	28.7
2003	27.0
2004	26.5
2005	25.6

* Number of licensed 16–17 year olds based on California Department of Motor Vehicles' annual driver license outstanding counts by gender, age and county. Age population data are based on U.S. Department of Census, 2000.

biases, no data or analyses have been produced to support this assertion, and it also ignores factors that would render the results conservative.

There is a second area of articulation between driver training and GDL. Some GDL programs allow drivers to achieve full licensure more quickly by passing an advanced driver training program. Evaluations of this time discount feature have found that those completing the training program have substantially higher subsequent crash rates than those not applying for the time discount (36,37). This does not mean, however, that the increased training has no effect on crash rates since one would expect large self-selection factors in determining whether a driver chooses to complete additional training in order to lessen the GDL restrictions. A valid analysis of the training program would require a design in which a pool of volunteers for the time discount were identified and randomly assigned to either training or no training conditions. This design would be difficult to execute without enabling legislation since it would require that those assigned to the control condition be given the time discount without receiving the training. What the above results do show is that any training effect is not sufficient to offset the effects of increased exposure, or decreased supervised driving – an outcome that should have been obvious on a priori grounds. The time discount may not be a good idea but it is important to understand what hypothesis is being tested. These studies are not evaluations of the causal effects of driver training on crash rates.

A third articulation between GDL and driver training concerns the effect on drivers who wait until they are adults to obtain their first drivers license, thereby bypassing the GDL process. In states like California, such drivers also avoid the mandatory driver training requirement. A study by Males (38) suggests that this may be leading to increased crash rates for 18–19 year olds but more research is needed to resolve this issue.

7. Conclusions

Despite the mixed quality of the studies reviewed and the variety of driver training programs evaluated, it is possible to offer some reasonable conclusions about the effectiveness of conventional driver training programs for novice drivers.

The most frequent characterization of the evidence encountered in the literature prior to GDL laws is that any effects are very small and short-lived and offset by the effect of training in accelerating licensure of 16–17 year olds. I believe this is a fair characterization of the evidence but it leaves unanswered to the question of “how small?” and whether the licensure offset results in a crash increase. I have argued that the evidence for a small positive effect on crashes per

licensed driver is more consistent with the data than an inference of a 0-effect. Admittedly, the primary basis for this inference is a single study (DeKalb) and an analysis with which some investigators have taken strong issue. My argument for a small positive non-zero effect is further buttressed by the fact that, if taken literally, all point null-hypotheses are false in the sense that two populations are never exactly the same on any measure. By very small effects, I doubt that a conventional driver training program reduces one-year crash rates by more than 5% – a value that is probably not cost-beneficial under most crash-cost models. As shown in Table 7, the sample size required to detect a 5% crash reduction using California data is prohibitive, probably exceeding the total sample sizes of all past studies combined.

Using a per capita crash metric, the evidence does suggest a net increase in crashes caused by driver training in a non-GDL state. However, I am not sure this justifies the conclusion that driver training has no positive effect, particularly if driver training increases the likelihood that a novice driver will pass the state licensing exam and have lower crash rates for the first 6–12 months following licensure. The conundrum raised by accelerating the licensure of a high risk group does raise legitimate public policy questions but is driver training really the core issue? A more obvious candidate is a state's minimum licensing age. If the crash rate of 16 year olds is deemed excessive due to immaturity, the most obvious and direct solution is to raise the minimum licensing age. This orientation is addressed in more detail in papers by McKnight (39) and Peck (40).

The issues addressed above change dramatically under GDL since the licensure rate problem vanishes. It therefore seems prudent that states which require driver training before granting a GDL retain that policy until future research evidence indicates otherwise.

One limitation with on-the-road training programs is that the primary focus is on skill; yet skill as measured by on-the-road tests has never been shown to be correlated with driver crash rates. In contrast, there have been numerous studies documenting the highly significant role of attitudinal and lifestyle factors in the high crash rate of young drivers (4,9,12,41,42). The role of attitudes was further documented in the DeKalb study, which found that the Mann driver attitude inventory was a much stronger predictor of crash rates than were road tests. This raises the difficult question of how to change the attitudinal and maturational factors underlying risky driving behavior through classroom and on-the-road training. It is difficult to see how simply requiring more hours of on-the-road training addresses the underlying problem (42).

Acknowledgements

I, the author would like to thank several individuals and colleagues who assisted me in locating papers and providing encouragement: Dr. Martin Lee-Gosselin, Scott Masten, Doug Luong, Dan Mayhew and Dr. Jim McKnight. I am indebted to my wife, Ellie Enriquez Peck, for typing this manuscript from handwritten drafts under a series of difficult circumstances.

References

- [1] K. Clinton, L. Lonero, Evaluation of Driver Education: Comprehensive Guidelines, AAA Foundation for Traffic Safety, Washington, DC, 2006.
- [2] J. Nichols, Driver education: a historical review, National Transportation Safety Board Public Forum on Driver Education and Training (Report of Proceedings), National Transportation Research Board, Washington, DC, 2003.
- [3] Association of Casualty and Surety Companies, Research on the accident reduction value of high school driver education, Traffic Safety Research Review 1 (2) (1957) 48–64.
- [4] D.M. Harrington, The young driver follow-up study: an evaluation of the role of human factors in the first four years of driving, Accident Analysis and Prevention 4 (1972) 191–240.
- [5] R.V. Rainey, J.J. Conger, C.R. Walsmuth, Personality characteristics as a selective factor in driver education, Highway Research Board Bulletin 285 (1961).

- [6] J.J. Conger, W.C. Miller, R.V. Rainey, Effects of driver education: the role of motivation, intelligence, social class and exposure, *Traffic Safety Research Review* 10 (1966) 67–71.
- [7] F.L. McGuire, R.C. Kersh, An Evaluation of Driver Education, University of California Press, Berkeley and Los Angeles, 1962.
- [8] J. Stock, J. Weaver, H. Ray, J. Brink, M. Sadof, Evaluation of safe performance secondary school driver education curriculum demonstration project, final report, U.S. Department of Transportation, National Highway Traffic Safety Administration, 1983.
- [9] R.C. Peck, The identification of multiple accident correlates in high risk drivers with specific emphasis on the role of age, experience and prior traffic violation frequency, *Alcohol, Drugs and Driving* 9 (1993) 145–166.
- [10] R.C. Peck, Novice driver training effectiveness evaluation, in: *Driver Education – Path Ahead*. Washington, DC: Transportation in Research Board Circular E-C 101. Paper presented at a conference sponsored by the TRB Committee on Operator Education and Regulation, 2006.
- [11] A. Lund, A. Williams, P. Zador, High school driver education: evaluation of the DeKalb County study, *Accident Analysis and Prevention* 18 (4) (1986) 349–357.
- [12] R. Jessor, S. Jessor, *Problem Behavior and Psychosocial Development*, Academic Press, New York, N.Y., 1977.
- [13] G.W. Milligan, D.S. Wong, P.A. Thompson, Robustness properties of non orthogonal analysis of variance, *Psychological Bulletin* 101 (3) (1987) 464–470.
- [14] D.J. DeYoung, An evaluation of the specific deterrent effect of vehicle impoundment on suspended, revoked and unlicensed drivers in California, *Accident Analysis and Prevention* 31 (1) (1999) 45–53.
- [15] B. Efron, D. Feldman, Compliance as an explanatory variable in clinical trials, *Journal of the American Statistical Association* 86 (413) (1991) 9–25.
- [16] N. Porta, C. Bonet, E. Cobo, Discordance between reported intention-to-treat and per protocol analyses, *Journal of Clinical Epidemiology* 60 (7) (2007) 663–669.
- [17] C.S. Davis, The DeKalb County George, driver education demonstration project: analysis of its long term effect, Washington, DC, U.S. Department of Transportation, National Highway Traffic Safety Administration, 1990.
- [18] L. Lonero, D. Mayhew, Large-Scale Evaluation Review of the Literature on Driver Education Evaluation – 2010 Update, AAA Foundation for Traffic Safety, Washington, DC, 2010.
- [19] P.R. Rosenbaum, D.R. Rubin, Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *American Statistician* 39 (1) (1985) 33–35.
- [20] R. Christie, The Effectiveness of Driver Training as a Road Safety Measure: a Review of the Literature, Royal Automobile Club of Victoria, Noble Park, Victoria, Australia, 2001.
- [21] D.R. Mayhew, H.M. Simpson, Effectiveness and Role of Driver Education and Training in a Graduated Licensing System, Traffic Injury Research Foundation, Ottawa, Ontario, 1996.
- [22] D. Mayhew, H. Simpson, The Safety Value of Driver Education and Training, in: B. Simon-Morton, J. Hartos (Eds.), *Proceedings of an Expert Conference on Young Drivers, Injury Prevention – Reducing Your Driver Crash Risk*, 8(2), 2002, pp. 3–8.
- [23] J. Woolley, In-car Driver Training at High Schools: a Literature Review, Safety strategy, transport A, Walkerville, South Australia, 2000.
- [24] L. Roberts, I. Kwan, School-based driver education for the prevention of traffic crashes, *Cochran Library 1*, Cochrane Data Base of Systematic Reviews, 3 (CD003201), Update software, Oxford, 2001.
- [25] J.S. Vernick, G. Li, S. Ogaitis, E.J. McKenzie, S.P. Baker, A.C. Gielen, Effects of high school driver education on motor vehicle crashes, violations and licensure, *American Journal of Preventive Medicine* 16 (1999) 40–46.
- [26] S.V. Masten, Teenage Driver Risks and Interventions (Report No. 04-207), California Department of Motor Vehicles, Sacramento, CA, 2004.
- [27] P.M. Strang, K.B. Deutsch, R.S. James, S.M. Mander, A Comparison of On-road and Off-road Driver Training, Victoria Road Safety, Hawthorne, Victoria Australia, 1982.
- [28] D.R. Dreyer, M.K. Janke, The effects of range vs. non-range driver training on accident and conviction frequencies of young drivers, *Accident Analysis and Prevention* 11 (3) (1979) 179–198.
- [29] K.J. Rathman, S. Greenland, *Modern Epidemiology*, 2nd ed. Lippincott-Raven Publishers, Philadelphia, PA, 1998.
- [30] D.T. Levy, Youth and traffic safety: the effects of driving age, experience and education, *Accident Analysis and Prevention* 22 (4) (1990) 327–334.
- [31] J.D. Wynne-Jones, P.M. Hurst, The AA driver training evaluation, Traffic Research Report No 33, Ministry of Transport, Road Transport Division, Wellington, New Zealand, 1984.
- [32] S.V. Masten, E.A. Chapman, The effectiveness of home-study driver education compared to classroom instruction: the impact on student knowledge and attitudes, *Traffic Injury Prevention* 5 (2) (2004) 117–121.
- [33] N.P. Gregersen, Systematic cooperation between driving schools and parents in driver education: an experiment, *Accident Analysis and Prevention* 26 (4) (1994) 453–461.
- [34] M. Janke, S. Masten, D. McKenzie, J. Gebers, S. Kelsey, Teen and Senior Drivers (Report No. 194), Department of Motor Vehicles, Sacramento, CA, 2003.
- [35] R. Elvik, T. Voa, *The Handbook of Road Safety Measures*, Elsevier, 2004.
- [36] A.F. Williams, M.A. Mayhew, Graduated licensing and beyond, *American Journal of Preventive Medicine* 35 (38) (2008) 324–333.
- [37] S. Wiggins, Graduated Licensing: Year Six Evaluation Report, Insurance Corporation of British Columbia, Victoria, Canada, 2006.
- [38] M. Males, California's graduated driver license law: effect on teenage deaths through 2005, *Journal of Safety Research* 38 (2007) 651–659.
- [39] A.J. McKnight, Driver education – when? in: D.R. Mayhew, H.M. Simpson, A.C. Donelson (Eds.), *Young driver accidents: In search of solutions*, Proceedings of an International Symposium, , 1985.
- [40] R.C. Peck, The effectiveness of novice driver education. Sacramento, CA: Department of Motor Vehicles (paper presented at Transportation Research Board Annual Meeting, January 8–11, 1996). Washington, DC.
- [41] A.F. Williams, Safety needs of novice drivers: lifestyle factors, in: *Driver Education at the Crossroads*. Washington, DC: Transportation Research Board Circular E-C024. Paper presented at a conference sponsored by the TRB Committee on Operation, Education and Regulation, 2001.
- [42] R.C. Peck, The role of youth in traffic accidents: a review of past and current California data, *Alcohol, Drugs and Driving* 1 (1985) 45–61.